DOCUMENT RESUME

ED 238 914 TM 840 004

AUTHOR	Boruch, Robert F.; And Others
TITLE	Recommendations to Congress and Their Rationale: The Holtzman Project.
INSTITUTION	Northwestern Univ., Evanston, Ill. Dept. of Psychology.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Feb 83
CONTRACT	OED-300-79-0467
GRANT	NIE-G-79-0128
NOTE	33p.; For related documents, see ED 192 466 and TM 840 006.
AVAILABLE FROM	Evaluation Review, 275 South Beverly Hills, Los Angeles, California 90212 (Publication No. 0193-841X/83/010005-31, $3.35).
PUB TYPE	Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT	Evaluation Review; v7 n1 p5-35 Feb 1983
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Education; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Utilization; Evaluators; *Federal Programs; Program Evaluation; School Districts; Standards; State Programs
IDENTIFIERS	Congress; *Holtzman Project; Secondary Analysis

ABSTRACT

This article concerns recommendations made in a report to Congress and the Department of Education on evaluation of federally supported education programs. The work covers local, state, and federal efforts to address questions about why and how well evaluations are done, and about how results are used. The recommendations are directed toward improving the quality of evaluations and enhancing their usefulness. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED238914

This article concerns recommendations made in a report to Congress and the Department of Education on evaluation of federally supported education programs. The work covers local, state, and federal efforts to address questions about why and how well evaluations are done, and about how results are used. The recommendations are directed toward improving the quality of evaluations and enhancing their usefulness.

RECOMMENDATIONS TO CONGRESS AND THEIR RATIONALE

The Holtzman Project

ROBERT F. BORUCH
DAVID S. CORDRAY
GEORGINE M. PION and LAURA C. LEVITON
Northwestern University

In June 1980, a group at Northwestern University produced a report for Congress on the evaluation of federally supported education programs at the national, state, and local levels. This article addresses only one aspect of the report—recommendations and the rationale for them.[1]

The report was undertaken in response to the Education Amendments of 1978 (Public Law 95-561). The relevant section of the law was introduced as a bill by Congresswoman Elizabeth Holtzman of New York, and it requires that the Secretary of Education conduct a comprehensive study of evaluation practices and procedures. The

AUTHORS' NOTE: *The research for this project was supported by OED-300-79-0467. The editorial work, follow-up studies on utilization, and recommendations have been supported by NIE-G-79-0128.*

EVALUATION REVIEW, Vol. 7 No. 1, February 1983 5-35
© 1983 Sage Publications, Inc.
0193-841X/83/010005-31$3.35

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

questions covered in the research are those implied by the law and the conference reports preceding it: Why and how are evaluations carried out? What are the capabilities of those who carry out evaluations? How are the results of evaluation used? What recommendations can be made to improve procedure or practice?

These questions were discussed with congressional staff and federal agency personnel to clarify them. The more detailed questions are elaborated in the body of the report. The Holtzman Project was prospective in its orientation, designed to provide evidence and argument bearing on these questions and to provide recommendations that will ameliorate the problems we identified. The project staff relied on two broad sources of information: contemporary investigations by other researchers and agencies, and direct field work. The latter included site visits to eight state education agencies (SEAs) and fourteen local education agencies (LEAs), and telephone surveys of approximately 200 LEAs. The site visits and the larger survey were based on a stratified random sample. Round-table discussions at Northwestern were undertaken to capitalize on experts in special topics, such as school board use of evaluation reports. Interviews with some staff members of all major federal agencies with an interest in educational evaluation (U.S. General Accounting Office—GAO, Congressional Budget Office—CBO, Congressional Research Service—CRS, and the executive operating units) were carried out. The literature review covered both unpublished and published documents, including reports maintained by ERIC and the LEXIS system. An earlier article, published in *Review of Educational Research*, served as a guide to reports issued before 1979.[2]

Recommendations to Congress are discussed first, and recommendations to the Department of Education (hereafter called "the Department") next. The rationale, qualifications, and limits to the recommendations follow.

RECOMMENDATIONS TO THE CONGRESS

PLANNING AND EXECUTING EVALUATIONS

We recommend that Congress direct the relevant staff of congressional committees, the GAO, and the CBO to meet regularly with

evaluation staff of the Department to (a) reach agreement about when particular evaluations are warranted, and the senses in which each evaluation required by law is possible; (b) clarify congressional information needs, quality of evidence required, and planning cycle for each major evaluation required by law; (c) identify specific committees and groups as audiences for evaluation results; (d) identify the changes in program or understanding that could result from alternative findings.

STATUTORY PROVISIONS FOR EVALUATION

We recommend that Congress, in constructing statutory provisions for evaluation, (a) specify exactly which questions ought to be addressed and the audiences to whom results should be addressed, when specification is feasible; (b) provide for formal assessment of the evaluability of the relevant program when specification is not possible; (c) provide for statistically valid field testing of proposed evaluation requirements when specification is not possible and in-house assessment is insufficient.

EVALUATOR CAPABILITIES

We recommend that (a) capabilities be assessed *before* new statutory evaluation requirements are directed to LEAs and SEAs to determine where resources are adequate to meet the demand; (b) training or technical assistance be expanded when the demands are notable and capabilities are low; (c) the feasibility and desirability of direct contract programs be explored to capitalize on LEA and SEA capabilities.

USE OF AND AUTHORITY FOR BETTER EVALUATION DESIGNS

We recommend that Congress (a) routinely consider pilot testing every major new program, major variations on existing programs, and major program components before they are adopted at the national level, using high quality evaluation designs; (b) authorize the Secretary of Education explicitly in each evaluation statute to use high quality designs (especially randomized field experiments) for planning and evaluating new program components, program variations, and new

programs, when estimates of program effect are desirable. We recognize that such estimates are not always appropriate or feasible.

CRITIQUE AND REANALYSIS OF EVALUATION RESULTS

We recommend that Congress, in statutory requirements for evaluation of major programs, (a) also require independent, balanced, and competent critique of evaluation results that are material to policy decisions; (b) require critique of *samples* of evaluations submitted by LEAs and SEAs in response to legal requirements; (c) require that statistical data produced by national evaluations be made available routinely for reanalysis.

USE OF EVALUATION RESULTS

We recommend that Congress (a) direct staff of relevant committees, the Department, and the GAO to routinely outline which institutions can reasonably be expected to use results of each major evaluation and how such results might be used, during the design stage of every major program evaluation; (b) specify exactly which evaluations have been used and why they were used, and which have not been used and why they were not used, in authorizations and appropriations committee reports; (c) require specific information about changes resulting from evaluation, whenever the law requires SEAs to describe uses of evaluation; (d) explore the feasibility of direct competitive grants and contracts programs focused on improving the use of results at the LEA and SEA levels.

STANDARDS AND GUIDELINES

Recently developed standards and guidelines for evaluation are not appropriate for incorporation into law. They are sufficiently well developed to recommend that Congress (a) use such guidelines to understand what can reasonably be expected of evaluations; (b) direct that state and federal agencies use them as a guide where appropriate to developing criteria for judging evaluation plans submitted by LEAs and SEAs; (c) elicit assistance in the interpretation of guidelines from congressional support agencies—such as GAO—that oversee the execution of policy, law, and regulations.

RECOMMENDATIONS TO THE DEPARTMENT OF EDUCATION

AUTHORITY FOR TECHNICAL DISCUSSION

We recommend that the Department authorize technical staff of evaluation units to initiate discussion of evaluation plans with pertinent congressional staff, at their discretion, and refrain from directives that impede direct discussion.

PLANNING AND EXECUTING EVALUATIONS

We recommend that the Department direct principal evaluation unit staff to meet regularly with relevant staff of committees to (a) negotiate agreement about when particular evaluations are warranted and the senses in which each evaluation required by law is possible; (b) clarify congressional information needs, quality of evidence required, and planning cycle for each major evaluation undertaken by the Department; (c) identify specific audiences or groups for evaluation results; (d) identify the changes in program or understanding that could occur on the basis of evaluation results.

TESTS OF NEW PROGRAM COMPONENTS, PROGRAM VARIATIONS, AND NEW PROGRAMS

We recommend that the Department authorize explicitly the use of high quality evaluation designs (especially randomized experiments when the interests lie in estimating program effects), in order to evaluate new program components; program variations; and new programs. The authorization should be incorporated into all regulations that require estimating the effects of innovative changes.

CRITIQUE AND SECONDARY ANALYSIS OF EVALUATION RESULTS

We recommend that the Department (a) provide for the independent, balanced, and competent critique of every major evaluation funded by the Department in procurement of evaluations and evaluation policy; (b) incorporate into procurement procedures and policy the requirement that all statistical data produced in major program evaluations be

documented and stored for secondary analysis; (c) create an administrative mechanism for deciding when simultaneous analysis by both the original evaluator and an independent analyst is desirable and feasible, and create a mechanism for executing simultaneous independent analyses.

ACCESS TO AND SPECIFICATION OF REPORTS

We recommend that the Department adopt a policy to (a) adhere to a clearance rule which makes evaluation reports available after a specified period of time; (b) specify completely the evaluation documents referred to in the Department's *Annual Evaluation Report*, the *Federal Register*, and policy statements; (c) include in every major evaluation report a list of core recipients of the report, or compile publicly available lists of core recipients.

THE USE OF EVALUATION RESULTS

We recommend that the Department direct evaluation unit staff or evaluation contractors to (a) on a regular basis, provide oral reports, as well as written reports, on results of major evaluations and on the uses to which results can be put, to relevant congressional staff and support agency staff and to the program staff within the Department; (b) create a system to periodically collect, synthesize, and report specific uses to which evaluations are put; (c) improve the *Annual Evaluation Report* by citing instances of use more specifically; (d) direct evaluation staff to meet regularly with congressional staff to clarify information needs, feasibility of evaluation, audiences for results, and ways in which results can be used to modify programs.

IMPLEMENTATION

We recommend that the Department (a) routinely require formal measurement of the degree to which program plans match actual operations; (b) adjoin research on methods of measuring implementation to the introduction of new programs and program variations; (c) create an inexpensive central information system on the time and resources required for full implementation of new programs.

RATIONALE

These recommendations are based partly on the project's findings *and* on our judgments about what needs to be done to improve evaluation practice. The latter—like all such judgments—is based on barely verifiable experience. We sought advice and criticism from some congressional and agency staff, members of the National Academy of Sciences Committee on Program Evaluation, and LEA and SEA staffs. We have capitalized on the professional forums identified earlier. Time, however, did not permit systematic critique. We are making as much information as possible available for competing analysis.

PLANNING AND EXECUTING EVALUATIONS

The legislative decision to evaluate is complicated by the large number of potential participants; namely, the Congress, the Department of Education's Office of Evaluation, the CBO, and the GAO. The time available to make the decision and to frame specific evaluation questions is variable and often appears to be insufficient. The advice of experts is only sometimes available. The process often leaves ambiguous the type of evaluation that is wanted, the audiences for the evaluation, the probable uses of the evaluation results, and the reasons an evaluation is wanted. The ambiguity leads to unnecessary squabbles and misdirected or belated efforts.

Elements for a More Orderly Process

The actions that appear to be necessary to improve matters include (a) regular meetings among evaluation staff of the Department and the pertinent congressional committees, (b) a planning system that matches evaluations to authorization cycles, (c) information systems that make access to previous work simpler and faster, and (d) identification of groups that can contribute to technical quality of the effort.

Reauthorization Cycles

There has been a recent effort to match the production of evaluations to the reauthorization cycle, and we understand from memos and recent

activity of the office of Assistant Secretary for Management that the effort will be sustained. It is imperative to do so if either management or Congress expects evaluations to be used in reauthorization decisions.

Meetings

There is no system of regular meetings among evaluation staff of the agency and the pertinent congressional committees in order to examine the senses in which a program can be evaluated. We believe that such meetings are essential to assuring that formal legislative demands for evaluation are as well informed as possible, and that the Department's evaluation unit is equipped to handle them. Ideally, such meetings should be held before the law requiring evaluation is enacted. If that is not possible, formal evaluability assessments should be undertaken as soon as possible after enactment.

Those meetings should focus on the information needs of Congress and the Department—notably on the questions that should be addressed in the evaluation. They should also be a vehicle for clarifying the reasons for asking the questions and identifying the audiences to whom answers ought to be addressed.

Apart from this, meetings might address chronic problems. Because different types of evaluation demand different resources, some agreement on salient issues needs to be made explicit, at least occasionally. The factors that influence the feasibility of evaluations—new versus old programs, for example—ought to be presented emphatically. Because every major evaluation must be tailored, the flexibility available, what is known and what is not known about the program, and so on ought to be made reasonably clear.

We do not mean to imply that a lockstep series of discussions among all relevant staff is warranted or possible. The point is that the absence of regular meetings on congressional needs virtually guarantees that some needs will not be met. That in turn invites buck-passing and evaluations of lower utility.

Relevant Groups

The groups that should be involved in the process include evaluation staff from the Department of Education's Office of Planning and Evaluation Services and from the pertinent congressional committees. It is

sensible to capitalize routinely on support agencies, such as the GAO's Institute for Program Evaluation and relevant divisions of the CBO and the CRS.

Interest Groups and their Role

Interest groups that draft bills which create or modify programs should be urged to also provide plans for evaluation of the program in the event that their suggestions are enacted as law. These plans should be routinely reviewed by the Department's evaluation unit, at least; at best, the review should include congressional staff, program staff, and *independent* critics

Impediments

There are impediments to any meetings of the sort proposed. Some agency staff members, for instance, have maintained that they have not been free to initiate conversations that would clarify the intent of a demand to evaluate, due to executive policy that restricts discussion. The restrictions are said to have a variety of legitimate origins, including preventing agency staff from lobbying directly and independently for pet programs, and to assure that there is at least some orderliness in dealing with the Congress. *For evaluation by units with authority to evaluate, however, we believe that such restrictions are inappropriate.* No evaluator can conscientiously address a question posed by Congress if the question cannot be discussed directly. We believe that agency policy must recognize the relative independence and discretion of evaluation units.

Impediments on the congressional-staff side appear to include committee staffers who will participate in no discussion unless directed by a committee chairman to do so. These appear to be in the minority. The more general problem is, we were told, time and the difficulty of coordinating meetings so as to be reasonably convenient to both agency staff and congressional staff.

STATUTORY PROVISIONS FOR EVALUATION

Though some statutes are specific about program reporting, references to evaluation in many statutes are very general. The simple require-

ment to evaluate the program or to evaluate the effectiveness of the program in meeting the objectives of the statute occurs frequently.

There is, however, great variety in the way individuals at the local, state, and federal levels of government interpret the word "evaluate" in law or elsewhere. It concerns the array of questions that might be addressed in an evaluation, in the approaches one might choose to answer them, and in the level of detail with which they might be answered.

More specific statement of the questions that need to be addressed can help to reduce confusion and ambiguity in what is intended by law, and it can facilitate understanding of the scope and probable costs and benefits of the information.

Specification

If Congress needs to know how many are served and how many are in need, what are services and their costs, what are the effects of programs on their primary or secondary clients, and what are the costs and benefits of alternatives, then Congress should request that information explicitly. The same discipline ought to be asked of interest groups, advisors, and others who draft evaluation language for programs. It can be requested that the questions for gauging effectiveness be specified along with other features of the program. That it is possible to be more specific is clear from the statute mandating the National Institute of Education (NIE) Compensatory Education Study. That such specification is *not* always sufficient is clear from the same study: Six months of discussion were needed after enactment to clarify evaluation goals.

It will not be possible or desirable to be explicit in every case. To assure that general demands for evaluation are not misinterpreted, the law should provide for a formal assessment of the senses in which the program can be evaluated within one year after the enactment of the legislation.

Need for Dialogue

Regardless of how well legal requirements can be specified, there is a persistent need for regular dialogue between agency staff and congressional staff in refining questions and developing agreements on what level of quality of evidence is warranted, and at what cost. The dialogue

has occasionally been encouraged in congressional committee reports, by some congressional staff and by some agency staff. But it is irregular and more heavily dependent on individual preferences than it should be.

Audiences

Because evaluation results may be directed to any number of audiences—Congress, Department management, interest groups, advisory committees, and so on—there is a clear need for focus. The more audiences there are, the more difficult evaluations become.

Pilot Tests of Evaluation Demands

When there is substantial disagreement about which questions should be addressed and about how the information might be used, pilot evaluations should be undertaken. That is, one mounts formal small-scale experiments to determine which of several different evaluation schemes work best. They can be put into the field (a) to measure the paperwork burden on respondents, (b) to determine the costs of collecting the information, (c) to determine the quality and usefulness of the information, and (d) to clarify language that can be used in statute and regulation.

USE AND AUTHORITY FOR BETTER EVALUATION DESIGNS

The authority to use better designs (especially randomized experiments) in the interests of less equivocal evaluations of new programs, major new program variations, and major new program components must be made explicit in law and regulation. This recommendation applies only when there is substantial interest in estimating the effects of a program on its primary target group.

By "randomized experiment," we mean assigning children, schools, or classrooms randomly to one of two or more program variations, for instance, and then observing their performance under each regimen. The random assignment guarantees that, in the long run, comparison of the variations will be fair. This is one of the reasons the design has been used in the Negative Income Tax Experiments, in the Manhattan Bail Bond experiments, in evaluation of television programs such as *Sesame*

Street, The Electric Company and *Free Style*, as well as in the evaluation of the effectiveness of medical treatments.

The rationale for the first part of the recommendation—pilot testing new programs—is that higher quality evaluations are more feasible before the program is adopted at the national level. Better evaluation designs can be employed, conclusions are less likely to be ambiguous, and political-institutional constraints are less likely to be severe. The introduction of new programs can be staged so that earlier stages constitute pilot tests for the later ones. This may seem terribly mundane to some readers. But recognize that in recent political discussion of the proposed Youth Incentives Program—an enterprise whose costs may exceed $850 million per year—there had been *no* formal attention to pilot testing or staged introduction before 1979. Title I compensatory education programs evolved in the same way ten years ago, and we still know pathetically little about effective variations. The simple notion that massive new programs ought to be pilot tested is still warranted.

The second part of the recommendation, concerning higher quality evaluation designs, is based on the presumption that we won't learn how to bring about detectable changes in the performance of children or schools without more conscientiously designed tests. The justification for the recommendation lies partly in the poor quality of designs used in the field. It is discouragingly easy to find, for example, legislative testimony in which a Title I program is declared to be a success by an individual because "test scores went up." We do *not* advocate attempting to estimate program effects in all cases. The process of estimating effects is complicated under the best of conditions. We advocate attention to high quality designs, especially randomized experiments.

At the local level, there are *some* evaluators with the interest and the skill to employ the design for the sake of fair tests. An obstacle, we believe, is confusion about authority for running such tests. So, for instance, an evaluator offered the opinion that the design is desirable, of course, but in the absence of a clear statutory mandate, that evaluator could not risk employing it. The failure of federal program managers to encourage randomized experiments at the local level is partly because the mandate to do so is not explicit. At the federal level, some authority already exists. Indeed, evaluations at that level (such as the one conducted for the Emergency School Assistance Act) have employed state-of-the-art experimental designs. But we believe that the authority must be made more explicit.

Feasibility

The usefulness of randomized tests, in principle, is not usually at issue in discussions about estimating effects of new programs. Argument about the uses of the design concerns the idea that randomized experiments are rarely feasible in field settings. Rarity does not establish lack of feasibility, and in any case a notable number of field tests have been mounted. (This is imperfect evidence, in that it doesn't guarantee that an experiment can be mounted successfully in the situation at hand.)

Pilot tests of experiments can yield more direct evidence on the feasibility of randomized experiments or other high quality designs. Consequently, we recommend mounting a small assessment prior to the major field experiments to identify problems in the field and to resolve them. Randomized experiments fail to be successfully implemented in education as in medicine, economics, and other fields, because the randomization is incomplete, because the programs are not implemented as advertised, and for other reasons. Pilot tests of the experiment itself can help to avoid unnecessary flaws in implementation.

Lacking dependable precedent and the opportunity for adequate pilot tests of the evaluation design, two general criteria for judging feasibility of randomized experiments are sensible. The first criterion hinges on the fundamental notion of equity. When there is an oversupply of eligible recipients for a scarce resource—program services—then randomized assignment of children to the resource seems fair. So, for instance, Vancouver's Crisis Intervention Program for youthful status offenders affords equal opportunity to eligible recipients. Since not all could be accommodated, and all are equally eligible, they are randomly assigned. Some experts argue that randomized experiments are most likely to be carried out successfully when the boon—real or imagined—is in short supply, and the demand for the boon is high. This rationale dovetails neatly with normal managerial constraints. That is, new programs cannot be emplaced all at once and not all eligible candidates can be served at once. Experiments can then be designed to capitalize on staged introduction of programs or services.

A second criterion concerns settings in which it is politically unacceptable to assign individuals randomly to control conditions,

despite the fact that we know nothing about whether a program works relative to no program at all. The ethical, moral, and economic justifications for experimenting may be quite irrelevant. In such cases, it is often possible to ameliorate difficulties by comparing program variations against one another, rather than comparing a novel program to an existing one or to no program at all. A "no-program" control condition may be an unacceptable political option, whether or not the program fails. The most we can reasonably expect then is to choose the invented variation component that works best for the investment.

The idea of testing variations or components, rather than testing a program against a control, is a compromise. But we believe that it is better than getting no information at all on the effects of the program in question. In particular, for ongoing programs that have strong public support, it seems sensible to think in terms of randomized assignment to new program variations or to new program components to discover more effective or cheaper versions of the program.

The most direct action that Congress can take to ameliorate the problem involves any statute that asks that the effects on children of a *new program variation*, or *new components* be estimated. We recommend that such statutes include an explicit provision authorizing statistically valid randomized experiments. For existing programs, we believe that some explicit authority is necessary to foster fair tests. That is, the Secretary of Education should be empowered to waive compliance with technical aspects of statutes or regulations for experimental projects that are likely to assist in promoting the statutory objectives. This would facilitate, for instance, randomized tests of cheaper variations on Title I programs, student loan programs, and the like.

INDEPENDENT CRITIQUE AND SECONDARY ANALYSIS

We recommend to Congress and the Department that major program evaluations be subjected routinely to competent, independent critique and secondary analysis. Mechanisms should be created to permit routine critique of a sample of evaluations produced at the LEA and SEA levels. By "critique" we do not mean adverse commentary. We do mean balanced examination of the quality of the report and judgments about whether recommendations can be sustained by the evidence. "Secondary analysis" here refers to analysis of raw statistical data, undertaken to improve on the quality of earlier analyses.

The origins of this recommendation lie partly in the idea that we should recognize good-quality evidence as such, and to properly identify poor evidence. There is also some need to prevent the ingenuous use of poor evidence and to avoid relying unnecessarily on one's confidence in the single evaluation. Furthermore, major evaluations are expensive. It seems sensible to allow the community of policy makers or their advisors to make the data work repeatedly, at low cost, in secondary analysis. Because evaluations may affect a variety of interest groups, those groups should be given an opportunity to offer competent criticism. Finally, we believe that the absence of independent criticism can reduce the importance of good evaluations.

Elements of a System for Critique and Secondary Analysis: National Level

The elements of an effective system for critique and secondary analysis include (a) explicit institutional policy on rapid disclosure of reports and access to statistical data underlying the reports; (b) a formal mechanism for independent critique or secondary analysis, when possible, *during* an evaluation; (c) a formal administrative mechanism for independent critique and secondary analysis when evaluation results are submitted; and (d) formal guidelines on reporting and storage of statistical information.

Elements of policy on reanalysis have already received some attention. For instance, the GAO has, in its guidelines on impact evaluation, taken the position that access to evaluative data for reanalysis is generally an important consideration. The Department of Education has not had a formal policy on disclosure of statistical data. However, the Department's Office of Planning and Evaluation has had an unwritten policy and has released data periodically for independent review and secondary analysis. Informal critiques of data sets have been undertaken by the CBO as a part of its efforts to screen studies for quality. These activities are undertaken so as to recognize individual privacy needs. Making policy formal, creating the administrative mechanisms, and testing them are sensible next steps.

Rapid access to evaluation reports has been a problem. Clearance of evaluation reports by the Secretary of Education, according to federal staff members who were interviewed, was slow at best. (We understand that the Department has recently adopted the 10-day clearance rule, which should improve matters.)

Informed Criticism

Opinions about the desirability of early independent review of major evaluations and of secondary analysis are not uniform. At least some agency staff reckon that a routine process will generate more heat than light. Assuring competent criticism in this arena is likely to be as difficult as it is in medicine, economics, and other fields. Some of that criticism is bound to be specious, dull witted, and self-interested. High quality in the design and execution of evaluations offers some protection against unwarranted criticism, but it is unlikely to be sufficient. Outcome evaluations are always subject to criticism, especially if the program does not work. We believe, however, that openness to criticisms must be given priority, and that some administrative research on reducing mindless criticism should be undertaken.

State and Local Level

A good many local and state evaluations provide no more than counts of those served, changes in test scores, and similar information. Regular, systematic reanalyses of the raw data underlying all of these is *not* warranted. It is more clear that samples of reports ought to be reviewed and criticized periodically. The main purpose of independent, competent criticism is to assure that the quality of evidence used to inform decisions is recognized. We also expect that this sort of critique will help to improve quality of the exercise, in the long run.

There are a variety of institutional vehicles available to conduct reviews. States with fairly well-developed evaluation units are a natural option. California and Michigan, for instance, have review, validation, and dissemination systems to try to assure that information about good programs of all kinds is available to LEAs. Such units may not, however, be independent of program offices. Moreover, some field investigation is warranted to determine whether or not evaluation capabilities are sufficient to generate high-quality critique.

The federal Joint Dissemination Review Panel (JDRP) is also a vehicle for critique of samples of evaluations. Its role is now limited to examining evidence volunteered by LEAs and other agencies that believe they are strong enough to sustain frank criticism, and so its mission would have to be expanded. The number of reviewers available on JDRP is not sufficient to review even a small additional sample, and so it would have to be enlarged

Technical Assistance Centers (TACs) supported under Title I constitute another option. But their role is confined to providing advice only when asked about Title I programs. It is not clear that TACs can be regarded as independent reviewers simply because they may provide advice on evaluations in the first instance.

The problems of assuring decent review and reanalysis of evaluation reports is sufficiently important to warrant further examination by federal agency management. That examination should address (a) alternative plans and administrative vehicles for critique and reanalysis, (b) alternative sample designs and time frame, and (c) design of pilot tests for review so as to estimate costs and benefits of a system before it is emplaced, to determine if the effort is indeed justified.

ACCESS TO REPORTS

Effective mechanisms to assure early release of evaluation reports and ready access to reports ought to be created. The origins of this recommendation lie partly in the idea that evaluation reports offered as a basis for policy, major executive decisions, and oversight should be open to competent criticism and should be accessible to a wide variety of potential users. The recommendation stems partly from the difficulty encountered in obtaining reports at the federal level, though at this level is far less difficult than it is at other levels of government.

Rapid access to reports prior to 1980 was impeded by clearance processes within the education division of the Department. That is, reports issued by a contractor have been reviewed by the Executive Secretariat of the Department before release, and those reviews have resulted in delays in release without notable improvement in the documents themselves.

The inclusion of a clause in Department contracts—Article 28—requiring that permission be sought prior to even discussing an evaluation, is more invidious. It prevents some universities from bidding on evaluations, since the clause runs counter to university standards of intellective independence. It is possible that this proviso reduces the quality of reports by impeding discussion of projects in professional forums. The mechanical difficulties of identifying and obtaining a report or a cluster of reports bearing on a specific evaluation are very tedious.

The problem of assuring rapid access has been rectified at least in the sense that the Office of the Deputy Assistant Secretary for Evaluation and Management has established a new clearance process. Reports are to be released automatically after 10 days if the secretary-level review has no modifications. The memorandum also permits adjoining criticism to the released document by program managers. We believe that automatic clearance after a specified period is desirable and that the practice ought to be maintained.

The practice of requiring contractors to seek permission for discussing results in public forums has not been changed, as far as we know. Our recommendation is that no such requirement be imposed in contracts.

Distribution of Information

We suggest the creation of a department-wide periodical that identifies and abstracts each evaluation report submitted to the Department and submitted by the Secretary of Education to Congress. We expect this to ameliorate access problems inside and outside the government. At its best, such a periodical will keep the public, Congress, and staff of the Department abreast of what has been produced and perhaps even of why it was produced. Models for this include the GAO's *Monthly Reports*, which summarizes reports issued by the agency.

Responsibility for Distribution

The practice of assigning sole responsibility to the project officer for final reports is not entirely effective. Officers vary in their attention to circulating reports and submitting them to distribution centers such as ERIC; they shift agencies, resign from government service, and otherwise disappear. So do reports, at least at times. Mechanisms must be developed to avoid reliance on the single officer. The options include (a) strengthening internal agency capability for storage of reports; (b) assuring that the list of core recipients for reports is included in the reports themselves, or that such a list is publicly available; (c) requiring the contractor and the agency to maintain a list of reports (with full citations) generated, together with the location of the agency that disseminates it; (d) requiring that the recipient of each evaluation executed under contract or grant provide reports, abstracts of reports, or both,

after 10-day clearance to ERIC, the National Technical Information Service (NTIS), the pertinent education centers and laboratories, the Committee on Evaluation and Information Systems (CEIS), the Federal Education Data Acquisition Council (FEDAC), and congressional staff and support agencies (especially the CRS and GAO); and (e) distributing each report routinely to every federal evaluation project officer and every evaluation contractor.

TRACKING THE USE OF EVALUATIONS

Our attention to this topic stems partly from the arguments about whether evaluations are used. The answer is easy: Some are and some are not. The more interesting questions concern how they are used, how often they are used, how to balance their cost against use and how to encourage use. The last two questions cannot be answered adequately now, because the hard answers to the "how" and "how often" questions are fragmentary, and the soft answers are rather too dependent on flawed memory and competing interests.

The problem of verifying use or nonuse hinges partly on turnover of staff responsible for initiating, conducting, and using evaluations. Corroborating use of an evaluation through independent sources is difficult and sometimes impossible. Titles of reports often imply nothing about potential or actual use, and reports are misremembered or forgotten. Incomplete citation is a chronic problem. The following recommendations are mundane but critical for inexpensive tracking. At best, they will eliminate part of the burden placed on respondents in studies of the use of evaluations.

Better Specification in Reports and Regulations

Congressional reports, agency annual reports, regulations, and the like are often imprudent or at least sloppy in failing to specify which evaluations they've used. Yet, such reports can be useful in tracking use of evaluations and in improving them. References to evaluations should then include author, title of the report, data of issuance, and sponsoring agency. If congressional or agency staff themselves cannot supply full references, then merely hiring an inexpensive, bright graduate student to build a specific reference list for each report of the half-dozen or so congressional committees most pertinent to educational evaluation

would suffice, so long as access to the list and dissemination of the list were assured.

The recommendation applies to both congressional committee reports and to major agency documents such as the Department's *Annual Evaluation Report* and policy statements. CBO documents are somewhat more conscientious, and GAO documents normally carry at least part of the information suggested. This recommendation also applies to proposed and final regulations issued in the *Federal Register*, since evaluations do result in regulation changes but are rarely recognized completely in the prose describing changes. An illustration of exemplary practice is the modification of regulations on day care in 1980.

The practice of recognizing evaluations explicitly when they have been useful in deliberations of Congress and at the executive level, is admirable. Identifying what is useful guides the agencies in the long run (if not the short run), rewards those who perform well, and exhibits some integrity to an occasionally cynical audience. The practice of recognizing good evaluations that are used is not uniform, however. The sponsoring agency is not given credit, due to time and resource constraints. That problem is serious enough to discourage some staff, if not to demoralize them. It would be helpful if more conscientious attention were given to recognizing useful evaluations and to recognizing *useless* evaluations in committee reports and the like.

Improving the Department's *Annual Evaluation Report*

The *Annual Evaluation Report* enumerates uses of evaluations completed by the Department. It is important, and there are some ways to make it more useful. (1) The report on use should provide specific citation of each evaluation report, its author, title, date of issue, and issuing agency. Otherwise, it's impossible for the reader to verify that a report has been issued, much less that it has been used. (2) The report on use should provide specific citation of hearings or congressional reports in which an evaluation report is mentioned or used, and specific citation of regulations that are said to have been changed on the basis of evaluation results. It should cite regulations that are proposed or created as a result of the evaluation. Otherwise, verifying claims of use is difficult or impossible. (3) The contributors to the section on use of evaluations should be acknowledged to permit verification and corroboration. (4) The *Annual Evaluation Report*'s perspective on use ought to be

reexamined to identify flaws in indicators of use, such as citations of hearings, and the possible biases in them. Ignoring agencies apart from Congress makes it likely that use of evaluation results is understated. Very little information on management uses, apart from regulations, is provided. (5) Evaluations for which it is difficult to find verifiable evidence of their use should be identified. Evaluations that are virtually useless two years after production should be identified explicitly. The issue that ought to be addressed in future examination of uses is, Would reporting other than annually make sense?

Identifying the Recipients of Reports

Major evaluation reports should include a list of the individuals, committees, and agencies to whom the report was sent. This will facilitate tracking the use or nonuse of reports and our understanding of misdirected effort. The practice of appending reader lists to reports is current at the Office of Naval Research. The practice appears to be feasible for at least major evaluation reports. Where enumeration of members of the audience is not feasible, then the lists used internally as a basis for distribution of reports ought to be accessible.

Tracking Management Changes

Very little systematic, publicly available evidence is available on the managerial uses of evaluation. Moreover, there is no general mechanism for regularly following up on whether problems identified in an evaluation have been or can be rectified. Follow-up does occur episodically, through questions addressed to managers at committee hearings for instance. But we have been unable to identify any special, orderly record-keeping on the matter. We recommend that a simple examination of alternative mechanisms be undertaken to determine if a cheap follow-up system can be developed, and to determine how such mechanisms can be field tested.

Local and State

We have not investigated state uses of evaluations sufficiently to make recommendations on tracking mechanisms at that level. However, two features of some local and state efforts are worth considering by

both federal and state agencies. Some states, such as Massachusetts and Michigan, require that in local reports to the state the various usages of evaluations be reported regularly. Those reports are, in principle, a vehicle for tracking use and, occasionally, synthesis. It is possible that a few states have developed especially efficient ways to accomplish this task. If so, the procedures ought to be made available to other state and federal agencies. The alternative to regular reporting is a special survey undertaken to obtain periodically a better picture of uses than one could obtain in reports. At least one state—California—has tried this option, and the results were informative.

STANDARDS AND GUIDELINES

Current guidelines can be exploited in designing evaluations and in making crude judgments about quality of an evaluation report; however, they are not equally relevant to all types of evaluation, and they are not appropriate for inclusion in law or regulation. They should be recognized in policy statements, internal guidelines, and other flexible directives.

Guidelines have been developed by the GAO, the Evaluation Research Society, and the independent Joint Committee on Standards for Educational Program Evaluation. Standards are embodied in manuals used by the federal Joint Dissemination Review Panel in assessing educational worth of new programs.

There is substantial overlap in topical coverage of all of these. Moreover, the topical coverage overlaps with standards used in choosing designs for major national evaluations and grants for evaluative work supported by NIE and the Department of Education.

The guidelines are very general, as any set of guidelines on completeness and quality of evidence must be, given the variety of forms that evaluation may take. It is sensible, for instance, to expect that an evaluation that purports to estimate a program's effects on children covers pertinent topics: evaluation design, source and quality of information, competing explanations, and so on. These elements are part of most good guidelines, but they are no substitute for training and judgment.

The main justification for recommending that guidelines be recognized is that we believe they can be useful in clarifying what is meant by quality of evaluation and in informing the public about what can

generally be expected of evaluation. Guidelines may also be of some assistance in protecting the competent evaluation from gratuitous criticism, and in identifying inept evaluation. They can be useful in reviewing proposals made by LEAs for programs that require special evaluation, such as bilingual education.

National Level

We recommend that guidelines be formally recognized as such by agency executives and by congressional committee staff. They have already been recognized by evaluation staffers within the education agencies and the GAO; indeed, agency staffers contributed to their development. By recognition we mean formally acknowledging their existence, assuring that pertinent staff know about them, and testing the guidelines in the field. It would not be difficult to incorporate short reviews of guidelines into training programs and seminars on evaluation run by the CRS, the GAO, or the Federal Executive Institute.

State and Local Level

It is reasonable to assure that SEAs and LEAs know about the guidelines, to make guidelines available, and to encourage tests of guidelines at the local level. Guidelines can, for instance, be cited in requests for proposals (RFPs) and grant material, without demanding that they be followed. They may be made available through special-purpose information clearinghouses, such as ERIC, or through commerical publishers.

It is reasonable to encourage their use, but not to *require* it, in the interests of fostering better quality evaluations and protecting competence. That encouragement can be given through federal and state agency offices that disburse funds for innovative programs.

Responsibility for advising the public, administrators, school boards, and so on currently rests with evaluation staff at local and state levels. It is not unreasonable to urge that they make guidelines available to these audiences for evaluation results. The guidelines are pertinent, however, to the minority of LEAs, namely, the ones that do more than simple monitoring.

Field Tests

We do not recommend incorporating guidelines into law or regulation. Only some aspects of guidelines have been field tested; and regardless of how reasonable they appear to be in principle, their costs and benefits need to be better established before they are generally required. It is also sensible to determine their susceptibility to incompetent interpretation, misinterpretation, and corruption. Finally, guidelines will change a bit as the state of the art in evaluation develops. Formal tests may help to avoid a prematurely rigid posture on what constitutes quality.

Caveats

Contemporary guidelines cannot be simply applied to evaluation reports produced by LEAs in response to federal or state reporting requirements. In the first place, reports differ appreciably in content, depending on audience. Reports made to Parent Advisory Committees in Title I programs, for instance, contain information that differs in depth and in kind from information provided to states. Second, federal requirements are minimal. Any review of what is produced to fill requirements is likely to be a useful target for guidelines, simply because reports are more useful at local and regional levels.

ESTIMATING THE EFFECT OF PROGRAMS

The general expectation that *all* local, state, and federal education agencies will produce clear evidence on the effects of programs should be abandoned. The emphasis should be placed on finding better variations on programs and effective program components in LEAs and SEAs that have the resources to plan fair field tests, and on well-designed tests run by the federal government.

Measuring growth of children in intellectual achievement, personal development, and other areas is often warranted. However, the practice of attributing growth to a program on the basis of these data alone is not warranted, simply because there are so many competing explanations for growth or any change. Local and state evaluations rarely recognize competing explanations.

The demand for information about how much a program affects children must recognize that clearly interpretable estimates of this depend on evaluation designs that accommodate competing explanations. Those designs are not always feasible in local settings. Technical assistance is no substitute for resources, for local interest in estimating effects, or for those designs. Moreover, estimating effects at the local level often has lower priority than providing services that children and their parents believe are effective.

The demand for estimates of effect on children induces a kind of benign hypocrisy among some staffers, administrators, and local contractors responsible for programs and evaluations. An increase in test scores is treated as evidence that the program "works," for instance. The conscientious members of each camp will admit that other explanations—normal growth, for instance—are possible. They also admit (and we agree) that separating out the influence of the program from other influences is not possible without a great deal of managerial, legal, and technical effort, and it may be impossible despite those efforts. The admission appears rarely in evaluation reports on Title I programs, vocational education, and bilingual education.

Judging from our site visits, some LEAs and SEAs are interested in testing cheaper varieties of programs, program components, and program variations, and some of these are capable of doing this well. It is sensible to capitalize on that interest, if the evaluations of these are well designed. To the extent possible, contracts for doing so ought to be made available. Funds have been available through Title IV-C and some NIE programs. They can lead to better understanding of what works, what works inexpensively, and to the dissemination of the programs to interested local agencies. The effort may have to be augmented with assistance from universities, private contractors, technical assistance centers, or others. These are not substitutes for in-house staff and for strong administrative support of fair tests from administrators and oversight groups.

The national interest in understanding effects of new programs, as well as the quality of delivery, needs to be recognized and reiterated. The conduct of pilot tests of new programs should be supported when feasible and appropriate. This recommendation stems partly from the progress made over the past ten years in mounting field tests of new programs, program variations, and program components. There have been imperfections and failures in these tests to be sure. The execution of

good outcome studies is exceedingly difficult. These problems should not be regarded as excuses to avoid the virtue of understanding effects. The public interest in evidence of this kind (in education or in other areas, such as medicine and economics) has not been consistent. Planned tests are always vulnerable on this account, as well as on account of their youth.

The questions about how money is spent and to whom services are delivered (so-called process evaluation or implementation studies) are also important. Judging by recent work, the emphasis on this information has been understated. There must be some stress, however, on obtaining more than body counts, to supply more than nominal statements of where dollars go and who receives the services. The character of services is often poorly understood. Any such investigation will not help one understand whether more notable effects are produced by certain services than by cheap competitors or by no service at all, of course.

ENCOURAGING INTEGRITY

Evaluation often engenders concern among those whose program is evaluated. This, in turn, can provoke institutional pressure to find nice results if the evaluator is under the supervision of the program manager. Consequently, maintaining integrity can be difficult. The following list of options was developed to understand how one might facilitate integrity in the face of such pressure at federal, state, and local levels.

Posture at the Policy, Management, and Oversight Levels of Government

There is some argument for the view that administrators of new and innovative projects should not be judged solely on the basis of the outcome of the program for which they are responsible. Many educational projects are high-risk ventures, and their failure is often, if not always, beyond the control of any individual or institution. It is important to understand why we fail. Program managers and their staffs, then, should be judged at least partly on the quality of evidence bearing on a program, regardless of whether one finds that the program itself is a success. To be effective, that view would have to prevail at national, state, and local levels.

Design of Evaluations of New Programs

It is sometimes possible to accommodate fear of evaluation through design of evaluations. One of several simple ways of doing so is not to evaluate program A by comparing it to no program at all. "No program at all" is often not a politically viable option if A fails. Rather, one ought to compare variation A of the program to variation B, where each variation has identical objectives, but differences in cost, approach, or other characteristics. The difficulty with this option is that we often lack the imagination or resources to invent B; and of course, it provides no information on the effects of A relative to no program at all.

External Review

One way to assure that incompetent evaluations and competent evaluations are properly labeled as such is to subject completed evaluations to external review. This tactic is consistent with the aims of the education agencies, the GAO, and other agencies with an interest in quality and standards of evidence. It is consistent with the recent trend toward secondary analysis of program evaluation data, conducted by independent academic institutions. The latter option has been used by, among others, the former U.S. Office of Education, the NIE, the Law Enforcement Assistance Administration, and other agencies in the United States. A variant on the tactic has been tried by individual researchers in Pakistan in reviewing evaluations sponsored by the government. This option cannot assure directly that evaluations done with integrity will be rewarded. Gratuitous criticism is common. It should make it more likely, however, that poor evaluations are recognized as such and are not rewarded.

Joint Dissemination and Review Approaches

Consider a review board that has clearly defined standards for examining the quality of evaluations, and which examines quality upon request from the program manager. A main objective of this panel or board is to officially verify the quality of the evidence and to declare that the program (if effective) deserves to be disseminated. Further, such a seal of approval can become a device for obtaining more money for similar projects from an agency. Both official recognition and the

opportunity to apply for dissemination funds are appreciated, we believe, by competent evaluators and program developers.

Such a system has been operating with some success within OE and NIE. The JDRP reviews educational products, basing review on evidence that conforms to articulated standards. Approval makes them eligible for money earmarked for expansion, dissemination, and other purposes.

Explicit Policy on Independence of Evaluation

There is no substitute at the national, state, or local level for policy on relative independence of the evaluator. Such policy can assure bureaucratic independence, notably eliminating clearance requirements for conversation or disclosure of reports to any group. It may involve administrative independence, notably by assuring that the evaluator report to an individual *other* than the program manager. It may involve fiscal independence, notably by assuring that funds earmarked for evaluation are channeled through the evaluation unit, by setting salaries for the unit independent of salaries for program operating units, and by other methods. It may involve political independence, for example, through the bipartisan approval of director of evaluation in the same spirit as appointments are approved for Inspector General and Comptroller General at the national level of government.

EVALUATOR CAPABILITIES

The primary reasons for suggesting that demands for evaluation be preceded by "capabilities assessment," particularly at the state and local levels, are discussed below.

First, identifying who is and who is not an evaluator (not to mention the appropriate competency level) is often difficult. Depending on the program and the assigned tasks, program staff, evaluation unit staff, outside contractors, or graduate students may have evaluation responsibility. Second, because the field is less than fifteen years old, few institutions offer formal certification in the area. There is considerable debate about training, and graduate curricula vary in emphasis across institutions.

More important, the skills and talents required of evaluators in LEAs and SEAs differ, depending on evaluation activity. When eval-

uation involves simply meeting minimal reporting requirements, the skills demanded do not require advanced graduate training, but some technical common sense is essential. When evaluation activities go beyond the minimum reporting requirements, the level and sophistication of required skills multiply quickly. These two types of activities and their capability demands should receive separate consideration in law, regulation, and evaluation policy.

By capabilities assessment we mean systematic attempt to describe the kinds of skills that are required for each kind of task. For national demands for evaluative information, this may involve intensive field research—task analyses—of good performers. It need not be elaborate, however. Observing what people do is better, but more expensive of course, than merely *asking* them what they do.

Meeting Federal Evaluation Reporting Requirements

It cannot be expected that *all* state and local education agencies have the capabilities necessary to adequately comply with federal evaluation reporting requirements. Often, program staff in these agencies – individuals with responsibilities other than evaluation—assume responsibility for reporting activities. These persons were not necessarily hired for their evaluation expertise. Consequently, technical assistance in evaluation should be provided so that agencies can adequately fulfill federal evaluation reporting requirements. It might be provided in a variety of ways. (1) At the minimum, the sponsoring agency should have direct access to evaluation unit staff with explicit responsibility for training in evaluation. These individuals can develop appropriate guidelines for evaluation, arrange evaluation workshops for individuals who must complete these requirements, and select the proper strategy for providing technical assistance. Federal program agencies without these resources should consider creating specific job positions in evaluation. (2) Adequate resources can be channeled to SEAs that administer these federal grant programs to permit them to provide easily accessible and expert technical assistance in evaluation. (3) Federally supported technical assistance centers, such as those existing for Title I evaluation, can be established to assist states and local education agencies in meeting federal reporting requirements. One approach is to expand services of the Title I TACs to provide evaluation assistance for other federal programs.

Technical assistance involves instruction and guidance in the actual conduct of evaluation; i.e., selection of program participants, use of tests, and completion of federal reporting forms. It also involves assistance in deciding who will evaluate. For example, districts that have capable evaluation units should be encouraged to use the services of the unit for all program evaluation needs. Small districts that do not have the resources to form their own research and evaluation unit may be instructed in other options, e.g., the formation of a consortium to hire competent evaluation staff who serve more than one district. Regional assistance centers can be developed or augmented in order to better provide technical assistance in evaluation. When outside contractors are employed, guidelines must be developed so that program staff and district selection boards can choose the most competent individuals, be sensitive to the types of skills required, and be aware of their rights in contractual arrangements. State guidelines of this kind are rare

Going Beyond Federal Evaluation Reporting Requirements

Some districts and states often attempt to go beyond federal reporting requirements. If competently executed, these evaluations can improve the quality of information submitted to federal agencies, Congress, and to such other audiences as Parent Advisory Councils and school boards. We believe that providing more opportunities to those LEAs and SEAs with interest and capabilities in evaluation is warranted. At the state level this can be accomplished through existing mechanisms, such as the monies targeted for improving state capabilities and state refinement grants for Title I evaluation supported by Section 183(c). These funding mechanisms should be supported. Dissemination of demonstrated improvements in evaluation practices developed by SEAs through contracts should be promoted.

The improvement of local education agency capabilities deserves more attention than it has received in terms of discretionary evaluation activities. While some of this can be accomplished through an expanded SEA role, other methods can be more specifically targeted at LEAs and supported directly from the federal government. One option is to expand the program of direct grants or contracts to LEAs for evaluation-related activities. This should allow LEAs to apply for and receive funds to engage in additional evaluation activities for federal programs or in research on ways to improve evaluation methods. A second option is to make available grants to LEAs/SEAs to foster

university LEA relationships. This might include funding for training programs jointly sponsored by academic institutions and LEAs/SEAs. This would not only provide training for agency personnel but would also improve the quality of evaluation programs in universities by allowing students to participate in actual evaluations. In addition, university-conducted workshops could be supported as an avenue of continuing education for education agency personnel. There may also be an opportunity to award matching monies for SEA/LEA investment in such arrangements as "an endowed chair," whereby university faculty can spend a period of time in these agencies conducting evaluations and designing procedures that will remain after their departure.

NOTES

1. The full report has been issued as R. F. Boruch and D.S. Cordray (eds.). *An appraisal of educational program evaluations: Federal, State, and local agencies.* Report to the Congress. Psychology Department, Northwestern University, Evanston, Illinois, 1980. It is available through the ERIC system, from the Office of Evaluation at the U.S. Department of Education, and the U.S. Government Printing Office (Document Number 1980 0-721-636/235). This article is a revision of Chapter 7 of the report. Revisions were based on reactions to presentations of this material (after the report was submitted) at the annual meeting of the Evaluation Research Society, the Northern Illinois Association for Research and Evaluation, the State of Illinois Department of Education, the CRS, the GAO, and elsewhere. The revisions are important, but they do not represent major deviations from the original text.

2. For brevity's sake, we omit references in this paper. See the full report for references. For a literature review, we refer you to R. F. Boruch and P. M. Wortman. Implications of educational evaluation for evaluation policy. *Review of Research in Education*, 1979, 7, 309-363.

Robert F. Boruch is Professor of Psychology and Education and Director of the Methodology Program at Northwestern University.

David S. Cordray is Assistant Professor of Psychology and Associate Director of the Methodology Program at Northwestern University.

Georgine M. Pion is Assistant Dean for Research at the University of Illinois.

Laura C. Leviton is a MacArthur Fellow at the University of Pittsburgh.

Publisher's Note: The following information is printed in accordance with U.S. postal regulations. Statement of Ownership, Management and Circulation (required by 39 U.S.C. 3685). 1. Title of Publication: EVALUATION REVIEW. 1A. Publication No.: 0193-841X. 2. Date of Filing: October 1, 1982. 3. Frequency of Issue: Bimonthly. 3A. No. of Issues Published Annually: 6. 3B. Annual Subscription Price: institutions, $62.00; individuals, $28.00. 4. Location of Known Office of Publication: 275 South Beverly Drive, Beverly Hills (Los Angeles), California 90212. 5. Location of the Headquarters or General Business Offices of the Publishers: 275 South Beverly Drive, Beverly Hills (Los Angeles), California 90212. 6. Names and Complete Addresses of Publisher, Editor, and Managing Editor: Publisher: Sara Miller McCune, 1102 South Oakhurst Drive, Los Angeles, California 90035. Editors: Richard A. Berk, University of California, Santa Barbara, California 93106, and Howard E. Freeman, Dept. of Sociology, University of California, Los Angeles, California 90024. Managing Editor: None. 7. Owner (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1% or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.): Sage Publications, Inc., 275 South Beverly Drive, Beverly Hills, California 90212; Sara Miller McCune, 1102 South Oakhurst Drive, Los Angeles, California 90035; George D. McCune, 1102 South Oakhurst Drive, Los Angeles, California 90035; Marvin Sirot, Mohawk Lane, Greenwich, Connecticut 06830; Gustave Sirot, Wingate Road, Guilford, Connecticut 06340; John B. Hurford, 220 East 63rd Street, New York, New York 10021; Albert L. Zesiger, 1125 Fifth Avenue, New York, New York 10028; Judith C. Zesiger, 1125 Fifth Avenue, New York, New York 10028; Morris Janowitz, 1357 E. 55th Place, Chicago, Illinois 60637. 8. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1% or More of Total Amount of Bonds, Mortgages or Other Securities: None. 9. For Completion by Nonprofit Organizations Authorized to Mail at Special Rates (Section 132.122, PSM): Not Applicable.

	Av. No. Copies Each Issue During Preceding 12 Months	Actual No. of Copies of Single Issue Published Nearest to Filing Date
10. Extent and Nature of Circulation		
A. Total no. copies printed (net press run)	3267	2625
B. Paid circulation		
1. Sales through dealers and carriers, street vendors and counter sales	0	0
2. Mail subscriptions	1837	607
C. Total paid circulation (sum of 10B1 and 10B2)	1837	607
D. Free distribution by mail, carrier or other means: samples, complimentary, and other free copies	106	98
E. Total distribution (sum of C & D)	1943	705
F. Copies not distributed		
1. Office use, left-over, unaccounted, spoiled after printing	1324	1920
2. Returns from news agents	0	0
G. Total (sum of E, F1 and 2—should equal net press run shown in A)	3267	2625

11. I certify that the statements made by me above are correct and complete. (Signed) Rhoda Blecker, Vice President.